Statistica Sinica **16**(2006), 135-151

ASYMPTOTIC RESULTS FOR FITTING SEMIPARAMETRIC TRANSFORMATION MODELS TO FAILURE TIME DATA FROM CASE-COHORT STUDIES

Lan Kong[1], Jianwen Cai[2] and Pranab K. Sen[2]

[1] *University of Pittsburgh and* [2] *University of North Carolina at Chapel Hill*

Abstract: Semiparametric transformation models are considered for failure time data from case-cohort studies, where the covariates are assembled only for a randomly selected subcohort from the entire cohort and additional cases outside the subcohort. We present the estimating procedures for the regression parameters and survival probability. The asymptotic properties of the resulting estimators are developed based on asymptotic results for U-statistics, martingales, stochastic processes and finite population sampling.

Key words and phrases: Asymptotic distribution, case-cohort design, consistency, finite population sampling, semiparametric transformation models, U-statistic.

1. Introduction

The cohort design is frequently advocated as superior to the case-control design in epidemiological studies because cohort design permits evaluation of absolute risk as well as relative risk, and gives an opportunity to study multiple outcomes related to a specific exposure. However, when the disease of interest is rare or the time between exposure and disease manifestation is very long, it is extremely costly to follow subjects until the occurrences of disease. Generally, most cost and effort involves the analysis of biological specimens or the ascertainment of covariate profiles from raw data. Prentice (1986) introduced a case-cohort design as a more efficient solution in large cohort studies and disease prevention trials. In a case-cohort design, expensive covariates are assembled only for a subcohort that is randomly selected from the entire cohort at the beginning of the study, and any additional cases/failures outside the subcohort.

Statistical methods for analyzing failure time data from case-cohort studies have been developed for the Cox hazards model and some alternative survival models. Here we consider a family of models, namely the semiparametric transformation models, for the case-cohort design where a subcohort is selected from the full cohort by a simple random sampling without replacement. Semiparametric transformation models incorporate a variety of non-proportional hazards models besides the Cox proportional hazards model and the proportional odds model. In the semiparametric transformation models, an unspecified strictly increasing function h of failure time T is linearly associated with a p-vector of covariates Z through the equation

$$h(T) = -Z'\beta + \varepsilon,$$

where Z' is a transpose of Z, β is an unknown p-vector of regression parameters and ε is a random error with a completely known distribution function F and density function f. In terms of survival function, semiparametric transformation models assume that a known transformation g of survival function $S_Z(t)$ is linearly associated with covariate vector Z by

$$g\{S_Z(t)\} = h(t) + Z'\beta,$$

where $g^{-1} = 1 - F$. In fact, the unspecified function h plays the same role as the baseline hazard function in the Cox proportional hazards model.

We recently proposed a weighted estimating equation method to analyze the case-cohort data with such models (Kong, Cai and Sen (2004)). The basic idea is to use the inverse probability weighting technique to extend the approaches of Cheng, Wei and Ying (1995, 1997) and Fine, Ying and Wei (1998) that are only valid for complete data. In this paper, we rigorously develop the asymptotic properties of the resulting estimators and explicitly state the sufficient conditions which have not been fully discussed in the previous research. The estimating procedures for regression coefficients and survival probability are briefly presented in Section 2. The corresponding asymptotic properties are stated and proven in Section 3.

2. Case-Cohort Estimators

2.1. Estimator of regression parameters

Let $\{T_i, C_i, Z_i\}$ $(i = 1, \ldots, N)$ be N independent, identically distributed (i.i.d.) copies of $\{T, C, Z\}$, where C is the potential censoring time. Assume that the distribution of C is independent of Z and T. Due to censoring, the observed data has the form (X_i, Δ_i, Z_i), where $X_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$ with $I(.)$ being an indicator function. For full cohort data, Fine et al. (1998) introduced an extra parameter $\zeta = h(t_0)$, where t_0 is a prespecified constant such that $pr\{\min(T, C) > t_0\} > 0$, and obtained the estimator of the parameter vector $\theta = (\zeta, \beta')'$ through the minimization of a sum of weighted squares

$$Q_w(\theta) = \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} w_{ij}(\hat{\theta}_u) \Big[\frac{\Delta_j I\{\min(X_i, t_0) \geq X_j\}}{\hat{G}^2(X_j)} - \eta_{ij}(\theta) \Big]^2,$$

where w_{ij} is a positive weight function for efficiency improvement, $\hat{\theta}_u$ is the unweighted least square estimator, $\hat{G}$ is the Kaplan-Meier estimator of survival function for censoring time, and $\eta_{ij}(\theta) = \int_{-\infty}^{\zeta}\{1-F(v+Z_i'\beta)\}dF(v+Z_j'\beta)$. The use of the truncation point t_0 is to ensure that $\hat{G}$ is uniformly consistent over $[0, t_0]$.

In a case-cohort design, suppose we select a subcohort of size n by simple random sampling without replacement from a cohort study that consists of N independent subjects. Each subject has the same probability $p = n/N$ to be selected into the subcohort. Let ξ_i be the subcohort indicator, taking value 1 if the ith subject is in the subcohort and zero otherwise, so $pr(\xi_i = 1) = p$. The failure status Δ_i is available for each subject. However, we only observe the data (X_i, Z_i) for the subjects in the subcohort ($\xi_i = 1$) and additional cases outside the subcohort ($\Delta_i = 1$ and $\xi_i = 0$). The conditional probability of observing the complete covariates for the i^{th} subject given the failure status is $\Delta_i + (1 - \Delta_i)p$. Motivated by the idea of weighting the incomplete data by the inverse selection probabilities (Horvitz and Thompson (1952)), we define a weight, namely ρ_{ij}, to reflect the contribution of a pair of subjects i and j to the estimating function. Specifically, $\rho_{ij} = \rho_i\rho_j$, where $\rho_i = \Delta_i + (1 - \Delta_i)\xi_i/p$. For estimating the parameter vector θ, we consider the weighted estimating function

$$U_N(\theta, \hat{G}) = \sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} \rho_{ij} w_{ij}(\theta)\dot{\eta}_{ij}(\theta)\left[\frac{\Delta_j I\{\min(X_i, t_0) \geq X_j\}}{\hat{G}_n^2(X_j)} - \eta_{ij}(\theta)\right], \quad (1)$$

where $\dot{\eta}_{ij}(\theta) = \partial\eta_{ij}(\theta)/\partial\theta$ and $\hat{G}_n$ is the Kaplan-Meier estimator of the survival function G for censoring time based on the subcohort data. We may obtain the estimator $\hat{\theta}$ by solving the equation $U_N(\theta, \hat{G}) = 0$. In the absence of censoring, $\eta_{ij}(\theta_0) = E[I\{\min(T_i, t_0) \geq T_j\}|Z_i, Z_j]$. Using the Kaplan-Meier estimate $\hat{G}_n$ to account for censoring data based on the inverse censoring probability weighting technique, the expectation of $\Delta_j I\{\min(X_i, t_0) \geq X_j\}/\hat{G}_n^2(X_j) - \eta_{ij}(\theta_0)$ tends to zero asymptotically. Thus, heuristically, (1) leads to an asymptotically unbiased estimating equation. The detailed proof is given in Section 3.

Let h_0 and β_0 be the true values of h and β respectively. Denote the marginal survival function of failure time by $R(t) = \int_z g^{-1}\{h_0(t) + Z'\beta_0\}dH(z)$, where H is the distribution of covariate vector Z. If $\tilde{s}$ is an index set of the subcohort, we can estimate the marginal cumulative hazard function Λ associated with failure time by $\hat{\Lambda}(t) = N^{-1}\sum_{i=1}^{N}\int_0^t\{\frac{1}{n}\sum_{l\in\tilde{s}} I(X_l \geq u)\}^{-1}dN_i(u)$. Also, noting that the conditional expected value $E\{I(X_i \geq t)/G(t)|Z_i\} = g^{-1}\{h_0(t) + Z_i'\beta_0\}$, we can obtain an estimator of $h_0(t)$, $\hat{h}(t)$, by solving the estimating equation

$V\{h(t)\} = 0$, where

$$V\{h(t)\} = \sum_{i=1}^{N} \Big[\rho_i g^{-1}\{h(t) + Z_i'\hat{\beta}\} - e^{-\hat{\Lambda}(t)}\Big].$$

Then the estimator of survival probability at a covariate vector z_0 is given by $\hat{S}_{z_0}(t) = g^{-1}\{\hat{h}(t) + z_0'\hat{\beta}\}$.

3. Asymptotic Results

Let s_0 and s_1 denote the index sets of all the censored observations and failures in the cohort, and let $\tilde{s}_0$ and $\tilde{s}_1$ denote the corresponding sets for the subcohort. Then the total subcohort set is $\tilde{s} = \tilde{s}_0 \cup \tilde{s}_1$, and the total cohort set $s = s_0 \cup s_1$. Let N_0 and n_0 be the numbers of censored observations in the cohort and subcohort respectively, and let N_1 and n_1 be the corresponding numbers of failures. Moreover, let $\mathcal{F}_0$ be the σ-algebra generated by $\{X_i, Z_i, \Delta_i = 0, i \in s_0\}$, $\mathcal{F}_1$ be the σ-algebra generated by $\{X_i, Z_i, \Delta_i = 1, i \in s_1\}$, and $\mathcal{F}$ be the σ-algebra generated by $\{X_i, Z_i, \Delta_i = 1, i \in s\}$. Denote $\sum_{i \neq j}^{m}$ as a double summation for $1 \leq i \neq j \leq m$, and similarly, $\sum_{i \neq j \neq k}^{m}$ as a triple summation for $1 \leq i \neq j \neq k \leq m$. Also, we take the martingale associated with the censoring time as $M_i^c(t) = I(\Delta_i = 0, X_i \leq t) - \int_0^t I(X_i \geq u) d\Lambda^c(u)$ for the i^{th} subject, where Λ^c is the common cumulative hazard function for censoring time. Naturally the conditions required for establishing the asymptotic distribution for the full cohort estimator in Fine et al. (1998) are also required for the case-cohort estimators. In addition, more conditions are necessary to ensure the desired asymptotic behavior of certain subcohort quantities. Specifically, we assume the following conditions for establishing the consistency and asymptotic normality of case-cohort estimators.

A. Covariate vector Z is in a compact set $\mathcal{L} \in R^p$.

B. $\mathrm{Var}\{w_{ij}(\theta_0)\dot{\eta}_{ij}(\theta_0)[\Delta_j I\{\min(X_i, t_0) \geq X_j\}/G^2(X_j) - \eta_{ij}(\theta_0)]\} > 0$, $\forall i, j = 1, \ldots, N$.

C. There exists a compact set Θ of $\theta_0 = (\zeta_0, \beta_0')'$ that satisfies
(i) Partial derivatives $\partial F(u - Z_i'\beta)/\partial u$, $\partial f(u - Z_i'\beta)/\partial u$ and $\partial^2 f(u - Z_i'\beta)/\partial u^2$ exist on $u \in (-\infty, \zeta)$ for all $i = 1, \ldots, N$, and they are uniformly continuous on Θ for any $Z \in \mathcal{L}$; and
(ii) $w_{ij}(\theta) > 0$ and $\partial w_{ij}(\theta)/\partial\theta$, denoted by $\dot{w}_{ij}(\theta)$, exists, w_{ij} and $\dot{w}_{ij}$ are uniformly continuous on Θ for all (i, j).

D. (i) $n/N \to \alpha$ $(0 < \alpha < 1)$ as $n, N \to \infty$; (ii) $N_0/N \to \nu$ $(0 < \nu < 1)$ as $N_0, N \to \infty$.

E. The limit of matrix $N^{-2} \sum_{i \neq j} w_{ij}(\theta_0)\dot{\eta}_{ij}(\theta_0)\dot{\eta}_{ij}'(\theta_0)$ exists as $N \to \infty$ and is positive definite.

F. (i) $\Lambda^c(t_0) < \infty$, and (ii) $\Lambda(t_0) < \infty$.

3.1. Asymptotic properties for $\hat{\theta}$

Three technical issues need to be carefully justified in the development of the asymptotic properties of the estimator $\hat{\theta}$. First, the simple random sampling of subcohort without replacement leads to lack of independence between the observations. Second, the sample size n_1 or n_0 is random, although the subcohort size n is fixed. Third, the estimating function itself is a variant of traditional U-statistic due to the use of the Kaplan-Meier estimator. By the Hoeffding decomposition, we first approximate the case-cohort quantity by the full cohort counterpart plus an additional term that is asymptotically uncorrelated to the full cohort part. Asymptotic results for the full cohort part are then readily established based on U-statistics theory, and the additional term can be handled by asymptotic results on finite population sampling. The martingale representation of the Kaplan-Meier estimator $\hat{G}_n$ is used to convert the estimating function to a traditional U-statistic. Before we show the consistency of $\hat{\theta}$, we provide some useful lemmas below. The corresponding proofs are given in the Appendix. Also, we consider the sample size n_0 as a fixed constant and delay the justification of its randomness to the end of this section.

Lemma 1. *Let $Y_1, \ldots, Y_N$ be i.i.d. random vectors where $Y_i = (X_i, \Delta_i, Z_i')'$, and let $\phi(Y_i, Y_j)$ be any vector function of Y_i and Y_j with $Var\{\phi(Y_i, Y_j)\} < \infty$. Define weight ρ_{ij} as in Section 2. Then under condition D, we have*

$$
\begin{aligned}
&N^{-\frac{3}{2}} \sum_{i \neq j}^{N} \rho_{ij} \phi(Y_i, Y_j) \\
&= N^{-\frac{3}{2}} \sum_{i \neq j}^{N} \phi(Y_i, Y_j) + \frac{\sqrt{N}}{n_0} \sum_{i \in \tilde{s}_0} \Big\{ \phi_N^*(Y_i) - \frac{1}{N_0} \sum_{k \in s_0} \phi_N^*(Y_k) \Big\} + o_p(1),
\end{aligned}
$$

where $\phi_N^*(Y_i) = [\nu^2/(N_0-1)] \sum_{j \in s_0 \backslash i} \{\phi(Y_i, Y_j) + \phi(Y_j, Y_i)\} + [\nu(1-\nu)/N_1]$ $\sum_{j \in s_1} \{\phi(Y_i, Y_j) + \phi(Y_j, Y_i)\}$. Furthermore, $N^{-2} \sum_{i \neq j}^{N} \rho_{ij} \phi(Y_i, Y_j) = N^{-2} \sum_{i \neq j}^{N}$ $\phi(Y_i, Y_j) + o_p(1)$.

Lemma 2. *Let $(\xi_1, \ldots, \xi_N)$ be a random vector containing n ones and $N - n$ zeros, with each permutation equally likely and $n/N \to \alpha \in (0, 1)$. Let X_i, $i = 1, \ldots, N$, be i.i.d. random variables. Let f and g be functions for which $N^{-1/2} \sum_{i=1}^{N} f(X_i)$ converges to a normal distribution $N(\mu_f, \sigma_f^2)$, and $N^{-1/2} \sum_{i=1}^{N}$ $g(X_i)$ converges to a normal distribution $N(0, \sigma_g^2)$. Let $G_N(X) = N^{-1} \sum_{i=1}^{N} g(X_i)$*

and $H_N(X,\xi) = n^{-1}\sum_{i=1}^N \xi_i f(X_i) - \bar{f}_N$, *where* $\bar{f}_N = N^{-1}\sum_{i=1}^N f(X_i)$. *Then*

$$\sqrt{N}\begin{pmatrix} G_N(X) \\ H_N(X,\xi) \end{pmatrix} \xrightarrow{D} BVN\left\{0, \begin{pmatrix} \sigma_g^2 & 0 \\ 0 & \alpha^{-1}(1-\alpha)\sigma_f^2 \end{pmatrix}\right\}. \tag{2}$$

Theorem 1. *Under conditions* (A)$-$(E), $\hat{\theta}$ *is a root-n consistent estimator of* θ_0.

Proof. The proof follows from the application of the Inverse Function Theorem, as in Foutz (1977). Under conditions A and C(i), $\dot{\eta}_{ij}(\theta)$ and $\ddot{\eta}_{ij}(\theta)$ exist and are uniformly continuous in $\theta \in \Theta$, with $\dot{\eta}_{ij}(\theta) = (1, Z_j')' \int_{-\infty}^{\zeta}\{1 - F(t + Z_i'\beta)\} df(t + Z_j'\beta) - (1, Z_i')' \int_{-\infty}^{\zeta} f(t + Z_i'\beta) dF(t + Z_j'\beta)$. This result, with condition C(ii), ensures that $\partial N^{-2}U_N(\theta, \hat{G})/\partial\theta$ exists and is uniformly continuous on Θ. If $d_{ij} = \Delta_j I\{\min(X_i, t_0) \geq X_j\}$,

$$\begin{aligned}
&\frac{\partial N^{-2}U_N(\theta,\hat{G})}{\partial\theta} \\
&= \frac{1}{N(N-1)}\sum_{i\neq j}^{N}\rho_{ij}\{\dot{w}_{ij}(\theta)\dot{\eta}_{ij}(\theta) + w_{ij}(\theta)\ddot{\eta}_{ij}(\theta)\}\Big\{\frac{d_{ij}}{\hat{G}_n^2(X_j)} - \eta_{ij}(\theta)\Big\} \\
&\quad - \frac{1}{N(N-1)}\sum_{i\neq j}^{N}\rho_{ij}w_{ij}(\theta)\dot{\eta}_{ij}(\theta)\dot{\eta}_{ij}'(\theta).
\end{aligned}$$

By a Taylor expansion of $N^{-2}U_N(\theta,\hat{G})$ around G,

$$N^{-2}U_N(\theta,\hat{G})$$

$$= \frac{1}{N(N-1)}\sum_{i\neq j}^{N}\rho_{ij}w_{ij}(\theta)\dot{\eta}_{ij}(\theta)\left\{\frac{d_{ij}}{G^2(X_j)} - \eta_{ij}(\theta)\right\} \tag{3}$$

$$- \frac{1}{N(N-1)}\sum_{i\neq j}^{N}\rho_{ij}w_{ij}(\theta)\dot{\eta}_{ij}(\theta)\left[\frac{2d_{ij}}{G^3(X_j)}\{\hat{G}_n(X_j) - G(X_j)\}\right] \tag{4}$$

$$+ o_p(N^{-\frac{1}{2}}).$$

From the uniform consistency of the Kaplan-Meier estimator on $t \in [0, t_0]$ (Fleming and Harrington (1991, p.115), and application of convergence results for finite population sampling, $\hat{G}_n(t)$ converges in probability to $G(t)$ uniformly in $t \in [0, t_0]$ under D(i). Moreover, $w_{ij}(\theta)$ and $\dot{\eta}_{ij}(\theta)$ are bounded on Θ due to condition C. Hence, (4) converges to zero almost surely as $n, N \to \infty$. Furthermore, we may apply Lemma 1 to (3) by setting $\phi(Y_i, Y_j) = w_{ij}(\theta)\dot{\eta}_{ij}(\theta)\{d_{ij}/G^2(X_j) - \eta_{ij}(\theta)\}$ under conditions B-D. As a result, $N^{-2}U_N(\theta,\hat{G})$ is asymptotically equivalent to

$\{N(N-1)\}^{-1}\sum_{i\neq j}^{N} w_{ij}(\theta)\dot{\eta}_{ij}(\theta)\{d_{ij}/G^2(X_j)-\eta_{ij}(\theta)\}$, which converges almost surely to $\int_{z_1,z_2} w_{12}(\theta)\dot{\eta}_{12}(\theta)\{\eta_{12}(\theta_0)-\eta_{12}(\theta)\}dH(z_1)dH(z_2)$ by the Strong Law of Large Numbers for U-statistics. It then follows that $N^{-2}U_N(\theta_0,\hat{G}) \xrightarrow{p} 0$. Similarly, we may show that $-N^{-2}\partial U_N(\theta,\hat{G})/\partial\theta$ converges pointwise in probability to a deterministic function $I(\theta)$ as $n, N\to\infty$, where $I(\theta)$ is equal to

$$-\int_{z_1,z_2}\Big[\{\dot{w}_{12}(\theta)\dot{\eta}_{12}(\theta)+w_{12}(\theta)\ddot{\eta}_{12}(\theta)\}\{\eta_{12}(\theta_0)-\eta_{12}(\theta)\}w_{12}(\theta)\dot{\eta}_{12}(\theta)\dot{\eta}_{12}'(\theta)\Big] \\ dH(z_1)dH(z_2).$$

Note that $-N^{-2}\partial U_N(\theta,\hat{G})/\partial\theta$ is itself a U-process indexed by θ with bounded kernel because $\ddot{\eta}_{ij}(\theta)$ and $\dot{w}_{ij}(\theta)$ are uniformly continuous on compact set Θ by condition C. It follows from the Uniform Law of Large Numbers for U-statistics with bounded kernel function (Sherman (1994)) that $-N^{-2}\partial U_N(\theta,\hat{G})/\partial\theta$ converges uniformly to $I(\theta)$ in $\theta\in\Theta$. Furthermore, $-N^{-2}\partial U_N(\theta,\hat{G})/\partial\theta|_{\theta=\theta_0}$ is positive definite by condition E. The assumptions of Theorem 2 in Foutz (1977) are verified and we may conclude that $\hat{\theta}$ is a root-n consistent estimator.

Theorem 2. *Let* $Y_i=(X_i,\Delta_i,Z_i')'$. *Under conditions* A, C *and* F(i), $N^{-3/2}U_N(\theta_0,\hat{G})$ *can be expressed as*

$$N^{-\frac{3}{2}}U_{N,f}(\theta_0)+\sqrt{N}\Big\{\frac{1}{n_0}\sum_{i\in\tilde{s}_0}\varphi_N(Y_i)-\frac{1}{N_0}\sum_{i\in s_0}\varphi_N(Y_i)\Big\}$$
$$+2\sqrt{N}\Big\{\frac{1}{n}\sum_{k\in\tilde{s}}\int_0^{t_0}\frac{q(t)}{\pi(t)}dM_k^c(t)-\frac{1}{N}\sum_{k=1}^{N}\int_0^{t_0}\frac{q(t)}{\pi(t)}dM_k^c(t)\Big\}+o_p(1),$$

where $$U_{N,f}(\theta_0)=\sum_{i\neq j}^{N}\psi(Y_i,Y_j)+2N\sum_{k=1}^{N}\int_0^{t_0}\frac{q(t)}{\pi(t)}dM_k^c(t),$$

$$\psi(Y_i,Y_j)=w_{ij}(\theta_0)\dot{\eta}_{ij}(\theta_0)\Big[\frac{\Delta_j I\{\min(X_i,t_0)\geq X_j\}}{G^2(X_j)}-\eta_{ij}(\theta_0)\Big],$$

$$\varphi_N(Y_i)=\frac{\nu^2}{N_0-1}\sum_{j\in s_0\backslash i}[\psi(Y_i,Y_j)+\psi(Y_j,Y_i)]$$
$$+\frac{\nu(1-\nu)}{N_1}\sum_{j\in s_1}[\psi(Y_i,Y_j)+\psi(Y_j,Y_i)],$$

$$a_{ij}(\theta_0)=w_{ij}(\theta_0)\dot{\eta}_{ij}(\theta_0)\frac{I\{\min(X_i,t_0)\geq X_j\}}{G^2(X_j)},\quad \pi(t)=\lim_{N\to\infty}\frac{1}{N}\sum_{l=1}^{N}I(X_l\geq t),$$

$$q(t)=\lim_{N\to\infty}\frac{1}{N(N-1)}\sum_{i\neq j}^{N}a_{ij}(\theta_0)\Delta_j I(X_j\geq t).$$

Proof. According to a Taylor series expansion of $U_N(\theta_0, \hat{G})$ around G,

$$N^{-\frac{3}{2}} U_N(\theta_0, \hat{G}) = N^{-\frac{3}{2}} U_N(\theta_0, G) + 2\sqrt{N} \frac{1}{N(N-1)} \sum_{i,j \in s, i \neq j}^{N} \rho_{ij} \psi_1(Y_i, Y_j) + o_p(1),$$

where $\psi_1(Y_i, Y_j) = w_{ij}(\theta)\dot{\eta}_{ij}(\theta)[(d_{ij}/G^3(X_j))\{G(X_j) - \hat{G}_n(X_j)\}]$. Using Gill's martingale representation of the Kaplan-Meier estimator (Gill (1980, p.37)),

$$\frac{G(X_j) - \hat{G}_n(X_j)}{G(X_j)} = \int_0^{X_j} \frac{I(Y_n(t) > 0)}{Y_n(t)} \sum_{k=1}^{n} dM_k^c(t)$$

with $Y_n(t) = \sum_{l \in \tilde{s}} I(X_l \geq t)$. We then may write

$$N^{-\frac{3}{2}} U_N(\theta_0, \hat{G}) = N^{-\frac{3}{2}} U_N(\theta_0, G) + \frac{2\sqrt{N}}{n} \sum_{k \in \tilde{s}} \int_0^{t_0} \frac{q_N(t)}{\pi_n(t)} dM_k^c(t) + o_p(1), \quad (5)$$

where $q_N(t) = [1/(N(N-1))] \sum_{i \neq j}^{N} \rho_{ij} w_{ij}(\theta_0)\dot{\eta}_{ij}(\theta_0)(d_{ij}/G^2(X_j))I\{X_j \geq t\}$ and $\pi_n(t) = (1/n)\sum_{l \in \tilde{s}} I(X_l \geq t)$.

The integral with respect to the martingale in (5) is no longer a martingale since the integrand is not a predictable process, but we show in the appendix that

$$\left\| \frac{1}{\sqrt{n}} \sum_{k \in \tilde{s}} \int_0^{t_0} \left\{ \frac{q_N(t)}{\pi_n(t)} - \frac{q(t)}{\pi(t)} \right\} dM_k^c(t) \right\| \xrightarrow{p} 0, \text{ as } n, N \to \infty. \quad (6)$$

Thus, we can further rewrite quantity (5) as

$$N^{-\frac{3}{2}} U_N(\theta_0, G) + \frac{2\sqrt{N}}{n} \sum_{k \in \tilde{s}} \int_0^{t_0} \frac{q(t)}{\pi(t)} dM_k^c(t) + o_p(1).$$

After the use of Lemma 1 for the first term $N^{-3/2}U_N(\theta_0, G)$ with $\phi(Y_i, Y_j) = \psi(Y_i, Y_j)$ and some simple algebra manipulations, we have Theorem 2.

Theorem 3. *$N^{-3/2}U(\theta_0, \hat{G})$ is asymptotically normally distributed with mean zero and variance matrix $\Sigma(\theta_0) = \Sigma_0(\theta_0) + \Delta(\theta_0)$, where*

$$\Sigma_0(\theta_0) = E\Big[\frac{1}{N^3} \sum_{i \neq j \neq k}^{N} \{\psi(Y_i, Y_j) + \psi(Y_j, Y_i)\}\{\psi(Y_i, Y_k) + \psi(Y_k, Y_i)\}'\Big]$$
$$-4 \int_0^{t_0} \frac{q(t)q'(t)}{\pi(t)} d\Lambda^c(t),$$
$$\Delta(\theta_0) = \frac{1-\alpha}{\alpha\nu} E\Big[\frac{1}{N_0} \sum_{i \in s_0} \{\varphi(Y_i) - \frac{1}{N_0} \sum_{k \in s_0} \varphi(Y_k)\}\{\varphi(Y_i) - \frac{1}{N_0} \sum_{k \in s_0} \varphi(Y_k)\}'\Big]$$
$$+4 \frac{1-\alpha}{\alpha} \int_0^{t_0} \frac{\{q(t) - \nu(1-\nu)q_{01}(t)\}q'(t)}{\pi(t)} d\Lambda^c(t),$$

and $q_{01}(t) = \lim_{N_0,N_1\to\infty}[1/(N_0N_1)]\sum_{i\in s_0}\sum_{j\in s_1} a_{ij}(\theta_0)\Delta_j I(X_j \geq t)$.

Proof. The result in Theorem 2 can be written as

$$N^{-\frac{3}{2}}U_N(\theta_0,\hat{G}) = N^{-\frac{3}{2}}U_{N,f}(\theta_0) + U_{n_0,N_0} + U_{n_1,N_1} + o_p(1), \text{ where}$$

$$U_{n_0,N_0} = \sqrt{N}\{\frac{1}{n_0}\sum_{i\in\tilde{s}_0}\varphi_N(Y_i) - \frac{1}{N_0}\sum_{i\in s_0}\varphi_N(Y_i)\}$$
$$+\frac{2n_0\sqrt{N}}{n}\Big\{\frac{1}{n_0}\sum_{i\in\tilde{s}_0}\int_0^{t_0}\frac{q(t)}{\pi(t)}dM_i^c(t) - \frac{1}{N_0}\sum_{i\in s_0}\int_0^{t_0}\frac{q(t)}{\pi(t)}dM_i^c(t)\Big\},$$
$$U_{n_1,N_1} = \frac{2n_1\sqrt{N}}{n}\Big\{\frac{1}{n_1}\sum_{k\in\tilde{s}_1}\int_0^{t_0}\frac{q(t)}{\pi(t)}dM_k^c(t) - \frac{1}{N_1}\sum_{k\in s_1}\int_0^{t_0}\frac{q(t)}{\pi(t)}dM_k^c(t)\Big\}.$$

The first term $N^{-3/2}U_{N,f}(\theta_0)$ is a U-statistic with kernel of degree two. It follows from the Central Limit Theorem for U-statistics that $N^{-3/2}U_{N,f}(\theta_0)$ converges to a zero mean normal distribution. Note that both U_{n_0,N_0} and U_{n_1,N_1} represent the difference in certain averages between the random sample and the corresponding population counterpart. Thus, by Hájek's Central Limit Theorem (1960) for finite population sampling, they are each asymptotically normal conditional on $\mathcal{F}_0$ and $\mathcal{F}_1$, respectively. Moreover, U_{n_0,N_0} is uncorrelated with U_{n_1,N_1} given fixed sample size n_0, and both U_{n_0,N_0} and U_{n_1,N_1} are uncorrelated with the full cohort quantity $U_{N,f}(\theta_0)$ by Lemma 2. As a result, $U_{N,f}(\theta_0)$, U_{n_0,N_0} and U_{n_1,N_1} are mutually independent and they jointly converge to a normally distributed random vector. Hence, $N^{-3/2}U_N(\theta_0,\hat{G})$ is asymptotically normal with mean zero. We omit the calculation of variance for brevity.

The matrix $\Sigma_0(\theta_0)$ is in fact the variance matrix corresponding to the full cohort counterpart, and matrix $\Delta(\theta_0)$ accounts for the extra variability due to the case-cohort design. By virtue of the Taylor expansion of $U_N(\hat{\theta},\hat{G})$ around θ_0 and the consistency of $\hat{\theta}$, we obtain the asymptotic distribution for estimator $\hat{\theta}$.

Theorem 4. $\sqrt{N}(\hat{\theta}-\theta_0)$ *is asymptotically normal with mean zero and variance matrix* $I^{-1}(\theta_0)\Sigma(\theta_0)I^{-1}(\theta_0)$, *where* $I(\theta_0) = \int_{z_1,z_2} w_{12}(\theta_0)\dot{\eta}_{12}(\theta_0)\dot{\eta}_{12}'(\theta_0)dH(z_1)$ $dH(z_2)$.

Remark. Let m denote the random sample size and M be the fixed sample size. It is known that the Central Limit Theorem holds for m whenever it holds for M if (i) $m/M \xrightarrow{p} C(>0)$ and (ii) Anscombe's (1952) condition holds. In the case-cohort design we considered, $M = n\nu$ and $m = n_0$, so $m/M \xrightarrow{p} 1(>0)$. Anscombe's condition is itself a by-product of the tightness part of the weak convergence of partial sum processes (Sen and Singer (1993, p.337)). Thus, the invariance principles for U-statistics given in Sen (1981) imply that the Anscombe's condition is satisfied for U-statistics based on finite population sampling. Now we may conclude that the asymptotic results stated previously remain in force when n_0 is random.

3.2. Asymptotic properties for $\hat{S}_{z_0}(t)$

We need two more lemmas.

Lemma 3. *Let A_i, $i = 1, \ldots, N$, be i.i.d. stochastic processes with nondecreasing sample paths, indexed by an interval $[0, \tau]$. If $EA_i^2(0) < \infty$ and $EA_i^2(\tau) < \infty$, then $(1/\sqrt{N}) \sum_{i=1}^N \{A_i - E(A_i)\}$ converges weakly in $l^\infty[0, \tau]$ to a tight Gaussian process.*

Lemma 4. *Let $(\xi_1, \ldots, \xi_N)$ be a random vector containing n ones and $N - n$ zeros, with each permutation equally likely. Let $A_1(t), \ldots, A_N(t)$ be i.i.d. random processes on $[0, \tau]$ with nondecreasing sample paths, where $EA_i^2(0) < \infty$ and $EA_i^2(\tau) < \infty$. Then $(1/\sqrt{N}) \sum_{i=1}^N \xi_i \{A_i - E(A_i)\}$ converges weakly in $l^\infty[0, \tau]$ to a tight Gaussian process.*

Lemma 3 is given as Example 2.11.16 in van der Vaart and Wellner (1996, p.215). Its proof relies on the bracketing central limit theorem. Lemma 4 is given as a proposition in Kulich and Lin (2000).

Theorem 5. *$\hat{S}_{z_0}(t)$ is a monotone function in t and is a uniformly consistent estimator on $t \in [0, t_0]$.*

Proof. Recall that $\hat{S}_{z_0}(t) = g^{-1}\{\hat{h}(t) + z_0'\hat{\beta}\}$. Since we have shown that $\hat{\beta}$ is a consistent estimator, it suffices to show that $\hat{h}(t)$ is monotone in t and is a uniformly consistent estimator. The estimating equation $V\{h(t)\} = 0$ implies that $N^{-1} \sum_{i=1}^N \rho_i g^{-1}\{\hat{h}(t) + Z_i'\hat{\beta}\} = e^{-\hat{\Lambda}(t)}$. Since $\hat{\Lambda}(t)$ is nondecreasing in t and $g^{-1} = 1 - F$ is nonincreasing, $\hat{h}(t)$ is nondecreasing in t. To show the uniform consistency of $\hat{h}(t)$, it suffices to show that $\hat{\Lambda}(t)$ is a consistent estimator of $\Lambda(t)$ on $t \in [0, t_0]$. Let $M_i(u) = N_i(u) - \int_0^u I(X_i \geq s) d\Lambda(s)$, $Y_n(u) = \sum_{j \in \tilde{s}} I(X_j \geq u)$, and $Y_N(u) = \sum_{j \in s} I(X_j \geq u)$. Then $\sqrt{N}\{\hat{\Lambda}(t) - \Lambda(t)\}$ is equal to

$$\int_0^t \frac{\frac{1}{\sqrt{N}} \sum_{i=1}^N dM_i(u)}{\pi_n(u)} - \sqrt{N} \int_0^t \frac{\{\frac{1}{n} Y_n(u) - \frac{1}{N} Y_N(u)\} d\Lambda(u)}{\pi_n(u)}. \tag{7}$$

By the Martingale Central Limit Theorem, $(1/\sqrt{N}) \sum_{i=1}^N dM_i(u)$ converges weakly to a tight Gaussian process under condition F(ii). Moreover, $N^{1/2} \int_0^t \{n^{-1} Y_n(u) - N^{-1} Y_N(u)\} d\Lambda(u)$ converges weakly to a tight Gaussian process by the Functional Central Limit Theorem, as we note that $N^{1/2}\{n^{-1}Y_n(u) - N^{-1}Y_N(u)\}$

converges weakly to a tight Gaussian process by Lemma 4 and $\Lambda(u)$ is bounded in $u \in [0, t_0]$. Therefore it follows, from arguments as in the proof of (6), that (7) is equal to

$$\int_0^t \frac{\frac{1}{\sqrt{N}} \sum_{i=1}^N dM_i(u)}{\pi(u)} - \sqrt{N} \int_0^t \frac{\{\frac{1}{n} Y_n(u) - \frac{1}{N} Y_N(u)\} d\Lambda(u)}{\pi(u)} + o_p(1),$$

which further implies that

$$\hat{\Lambda}(t) - \Lambda(t) = \int_0^t \frac{\frac{1}{N} \sum_{i=1}^N dM_i(u)}{\pi(u)} \tag{8}$$

$$- \int_0^t \frac{\{\frac{1}{n} Y_n(u) - \frac{1}{N} Y_N(u)\} d\Lambda(u)}{\pi(u)} + o_p(\frac{1}{\sqrt{N}}). \tag{9}$$

The term (8) converges to zero in probability due to the martingale property. The term (9) is a difference between the subcohort average and cohort average, it also converges to zero in probability because $n^{-1} Y_n(u) - N^{-1} Y_N(u) \xrightarrow{p} 0$ uniformly in u, $\Lambda(u)$ is bounded, and $\pi(u)$ is bounded away from zero on $u \in [0, t_0]$. Thus, $\hat{\Lambda}(t)$ converges pointwise to true $\Lambda(t)$. Furthermore, this convergence is uniform in $t \in [0, t_0]$ because $\hat{\Lambda}(t)$ is monotone and bounded, and $\Lambda(t)$ is bounded and continuous. Consequently, we have that $e^{-\hat{\Lambda}(t)} \xrightarrow{p} e^{-\Lambda(t)}$. Using the large sample property of finite population sampling, we may show that as $n_0, N_0 \to \infty$,

$$\frac{1}{N} \sum_{i=1}^N \rho_i g^{-1}\{\hat{h}(t) + Z_i' \hat{\beta}\} - \frac{1}{N} \sum_{i=1}^N g^{-1}\{\hat{h}(t) + Z_i' \hat{\beta}\} \xrightarrow{p} 0. \tag{10}$$

In addition, $N^{-1} \sum_{i=1}^N g^{-1}\{h_0(t) + Z_i' \beta_0\} \xrightarrow{p} e^{-\Lambda(t)}$ by the Law of Large Numbers. Thus,

$$\frac{1}{N} \sum_{i=1}^N g^{-1}\{\hat{h}(t) + Z_i' \hat{\beta}\} - \frac{1}{N} \sum_{i=1}^N g^{-1}\{h_0(t) + Z_i' \beta_0\} \xrightarrow{p} 0, \text{ as } n_0, N_0, N \to \infty.$$

This implies that $\hat{h}(t)$ uniformly converges to $h_0(t)$ in $t \in [0, t_0]$ because $\hat{\beta}$ is a consistent estimator, and $\hat{h}(t)$ is monotone and bounded.

Theorem 6. $W_{z_0}(t) = \sqrt{N}[g\{\hat{S}_{z_0}(t)\} - g\{S_{z_0}(t)\}]$ *converges weakly in* $l^\infty[0, t_0]$ *to a Gaussian process with zero drift.*

Proof. A Taylor expansion of $V\{\hat{h}(t)\}$ around $h_0(t)$ and β_0 yields

$$\begin{aligned}
&N^{-\frac{1}{2}}V\{\hat{h}(t)\}\\
&= N^{-\frac{1}{2}}V\{h_0(t)\} - N^{\frac{1}{2}}\{\hat{h}(t)-h_0(t)\}N^{-1}\sum_{i=1}^{N}\rho_i f\{h_0(t)+Z_i'\hat{\beta}\} + o_p(1)\\
&= N^{-\frac{1}{2}}\sum_{i=1}^{N}\Big[\rho_i S_{z_i}(t) - \rho_i f\{h_0(t)+Z_i'\beta_0\}Z_i'(\hat{\beta}-\beta_0) - e^{-\hat{\Lambda}(t)}\Big]\\
&\quad -N^{-1}\sum_{i=1}^{N}\rho_i f\{h_0(t)+Z_i'\beta_0\}N^{\frac{1}{2}}\{\hat{h}(t)-h_0(t)\} + o_p(1).
\end{aligned}$$

Similar to the result in (10), we have that

$$N^{-1}\sum_{i=1}^{N}\rho_i f\{h_0(t)+Z_i'\beta_0\} \to a(t) = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} f\{h_0(t)+Z_i'\beta_0\} \text{ and}$$

$-N^{-1}\sum_{i=1}^{N}\rho_i Z_i f\{h_0(t)+Z_i'\beta_0\} \to b(t) = -\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} Z_i f\{h_0(t)+Z_i'\beta_0\}$ in probability as $n_0, N_0 \to \infty$. Therefore, it follows from $V\{\hat{h}(t)\} = 0$ and Taylor expansion of $e^{-\hat{\Lambda}(t)}$ around $\Lambda(t)$ that

$$\begin{aligned}
a(t)N^{\frac{1}{2}}\{\hat{h}(t)-h_0(t)\} &= N^{-\frac{1}{2}}\sum_{i=1}^{N}\{\rho_i S_{z_i}(t)\} + b'(t)N^{\frac{1}{2}}(\hat{\beta}-\beta_0)\\
&\quad -N^{\frac{1}{2}}e^{-\Lambda(t)} + N^{\frac{1}{2}}e^{-\Lambda(t)}\{\hat{\Lambda}(t)-\Lambda(t)\} + o_p(1).
\end{aligned}$$

As previously shown, $N^{1/2}(\hat{\theta}-\theta)$ is equivalent to $N^{-3/2}I^{-1}(\theta_0)U_N(\theta_0)$, so $N^{1/2}$ $(\hat{\beta}-\beta_0)$ is equivalent to $N^{-3/2}H(\theta_0)U_N(\theta_0)$, where $H(\theta_0)$ is obtained by removing the first row from the matrix $I^{-1}(\theta_0)$. It then follows from some algebraic manipulation that

$$\begin{aligned}
W_{z_0}(t) &= \sqrt{N}[g\{\hat{S}_{z_0}(t)\} - g\{S_{z_0}(t)\}] = \sqrt{N}[\hat{h}(t)-h_0(t)+z_0'(\hat{\beta}-\beta_0)]\\
&= \frac{1}{a(t)}\Big\{[b(t)+a(t)z_0]'H(\theta_0)N^{-\frac{3}{2}}U_N(\theta_0)\Big\} + W_1(t) + W_2(t) + o_p(1),
\end{aligned}$$

where $\quad W_1(t) = \dfrac{1}{a(t)}\{U_N^{(1)}(t) + \nu U_{n_0,N_0}^{(1)}(t)\}$,

$$W_2(t) = \frac{S(t)}{a(t)}\Big\{U_N^{(2)}(t) - \int_0^t \frac{U_{n,N}^{(2)}(u)}{\pi(u)}d\Lambda(u)\Big\},$$

$$U_N^{(1)}(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{S_{z_i}(t) - S(t)\}, \quad S(t) = \exp\{-\Lambda(t)\},$$
$$U_{n_0,N_0}^{(1)}(t) = \sqrt{N}\{\frac{1}{n_0} \sum_{i\in\tilde{s}_0} S_{z_i}(t) - \frac{1}{N_0} \sum_{i\in s_0} S_{z_i}(t)\},$$
$$U_N^{(2)}(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \int_0^t \frac{1}{\pi(u)} dM_i(u),$$
$$U_{n,N}^{(2)}(u) = \sqrt{N}\{\frac{1}{n} Y_n(u) - \frac{1}{N} Y_N(u)\}.$$

By similar arguments as in the proof of Theorem 3, $W_{z_0}(t)$ converges to a normal distribution at given t. The finite-dimensional distribution convergence is satisfied by Crámer-Wold device. To prove the tightness of W_{z_0}, it suffices to show that $W_1(t)$ and $W_2(t)$ are tight because $a(t)$ and $b(t)$ are not random. Application of Lemma 3 implies that the process $U_N^{(1)}(t)$ is tight because $S_{z_i}(t)$ is a monotone process and is bounded on $t \in [0, t_0]$. The process $U_N^{(2)}(t)$ is a martingale and the finite-dimensional distribution convergence of martingale ensures tightness (Sen (1981), Corollary 2.4.4.1). By virtue of Lemma 4, $U_{n_0,N_0}^{(1)}(t)$ and $U_{n,N}^{(2)}(t)$ each converge to a tight Gaussian process conditional on $\mathcal{F}_0$ and $\mathcal{F}$, respectively. Also, $U_{n_0,N_0}^{(1)}(t)$ and $U_{n,N}^{(2)}(t)$ are tight unconditionally by similar arguments as in the proof of Lemma 2 and characteristic function methods. Noting that function $\Lambda(.)$ is absolutely continuous with respect to Lebesgue measure, and the fact that linear functional of the Gaussian process is Gaussian, $\int_0^t U_{n,N}^{(2)}(u)\{\pi(u)\}^{-1} d\Lambda(u)$ is tight too. Now we may conclude that $W_{z_0}(t)$ converges weakly in $l^\infty[0, t_0]$ to a Gaussian process.

Acknowledgements

The authors are grateful to the reviewers for their constructive suggestions that have led to considerable improvement of the earlier version. This research was partly supported by National Institutes of Health NHLBI Grant R01 HL-57444.

Appendix

Proof of Lemma 1. Since $n_0/N \to \nu\alpha$ as $n_0, N \to \infty$, and $N^{-2} - N^{-1}(N -$

$1)^{-1} \to 0$ as $N \to \infty$, it suffices to show that

$$\frac{1}{N(N-1)} \sum_{i \neq j}^{N} \rho_{ij} \phi(Y_i, Y_j) \tag{11}$$
$$= \frac{1}{N(N-1)} \sum_{i \neq j}^{N} \phi(Y_i, Y_j) + \frac{1}{n_0} \sum_{i \in \tilde{s}_0} \{\phi_N^*(Y_i) - \frac{1}{N_0} \sum_{k \in s_0} \phi_N^*(Y_k)\} + o_p(n_0^{-\frac{1}{2}}).$$

Substituting the realization of ρ_{ij} into (11) yields

$$\frac{1}{N(N-1)} \Big[\frac{1}{p} \sum_{i \in \tilde{s}_0} \sum_{j \in s_1} \{\phi(Y_i, Y_j) + \phi(Y_j, Y_i)\} + \frac{1}{p^2} \sum_{i,j \in \tilde{s}_0, i \neq j} \phi(Y_i, Y_j)$$
$$+ \sum_{i,j \in s_1, i \neq j} \phi(Y_i, Y_j) \Big].$$

The second summation term is related to a U-statistic based on a simple random sample of size n_0, namely $U_{n_0,n_0} = n_0^{-1}(n_0-1)^{-1} \sum_{i,j \in \tilde{s}_0, i \neq j} \phi(Y_i, Y_j)$. The corresponding symmetric kernel is $\phi^+(Y_i, Y_j) = \{\phi(Y_i, Y_j) + \phi(Y_j, Y_i)\}/2$. Define the population counterpart of U_{n_0,n_0}, $U_{N_0,N_0} = N_0^{-1}(N_0-1)^{-1} \sum_{i,j \in s_0, i \neq j} \phi(Y_i, Y_j)$. Under the conditions that $\mathrm{Var}\{\phi(Y_i, Y_j)\} < \infty$, $n/N \to \alpha$ and $N_0/N \to \nu$ as $n, N_0, N \to \infty$, the Hoeffding Decomposition of U_{n_0,n_0} yields

$$U_{n_0,n_0} = \frac{2}{n_0} \sum_{i=1}^{n_0} [\phi_1^*(Y_i) - U_{N_0,N_0}] + U_{N_0,N_0} + o_p(n_0^{-\frac{1}{2}})$$
$$= \frac{2}{n_0(N_0-1)} \sum_{i \in \tilde{s}_0} \sum_{j \in s_0 \backslash i} \{\phi^+(Y_i, Y_j)\} - U_{N_0,N_0} + o_p(n_0^{-\frac{1}{2}}),$$

where $\phi_1^*(Y_i) = E[\phi^+(Y_i, Y_j)|Y_i] = (N_0-1)^{-1} \sum_{j \in s_0 \backslash i} \phi^+(Y_i, Y_j)$. Furthermore, with $U_{N,N} = N^{-1}(N-1)^{-1} \sum_{i \neq j}^{N} \phi(Y_i, Y_j)$, (11) is equal to

$$U_{N,N} - \frac{2}{N(N-1)} \Big[\sum_{i \neq j, i,j \in s_0} \{\phi^+(Y_i, Y_j)\} + \sum_{i \in s_0} \sum_{j \in s_1} \{\phi^+(Y_i, Y_j)\} \Big]$$
$$+ \frac{2}{N(N-1)p} \sum_{i \in \tilde{s}_0} \sum_{j \in s \backslash i} \{\phi^+(Y_i, Y_j)\} + o_p(n_0^{-\frac{1}{2}})$$
$$= U_{N,N} + \frac{2}{N(N-1)} \Big[\frac{1}{\alpha} \sum_{i \in \tilde{s}_0} \sum_{j \in s \backslash i} \{\phi^+(Y_i, Y_j)\} - \sum_{i \in s_0} \sum_{j \in s \backslash i} \{\phi^+(Y_i, Y_j)\} \Big]$$
$$+ o_p(n_0^{-\frac{1}{2}}).$$

Define $\phi_{n_0,N_0}(Y_i) = 2(N_0-1)^{-1} \sum_{j \in s_0 \backslash i} \{\phi^+(Y_i, Y_j)\}$ and $\phi_{n_0,N_1}(Y_i) = 2N_1^{-1} \sum_{j \in s_1} \{\phi^+(Y_i, Y_j)\}$ for $i \in \tilde{s}_0$. Then $E\{\phi_{n_0,N_0}(Y_i)|\mathcal{F}_0\} = 2U_{N_0,N_0}$ and $E\{\phi_{n_0,N_1}$

$(Y_i)|\mathcal{F}_0\} = 2(N_0N_1)^{-1}\sum_{i\in s_0}\sum_{j\in s_1}\{\phi^+(Y_i,Y_j)\}$. Now we rewrite $N^{-1}(N-1)^{-1}\sum_{i\neq j}^N \rho_{ij}\phi(Y_i,Y_j)$ as

$$
\begin{aligned}
& U_{N,N} + \frac{1}{N(N-1)}\Big\{N_0(N_0-1)\frac{1}{n_0}\sum_{i\in\tilde{s}_0}[\phi_{n_0,N_0}(Y_i) - E\{\phi_{n_0,N_0}(Y_i)|\mathcal{F}_0\}] \\
& +N_0N_1\frac{1}{n_0}\sum_{i\in\tilde{s}_0}[\phi_{n_0,N_1}(Y_i) - E\{\phi_{n_0,N_1}(Y_i)|\mathcal{F}_0\}]\Big\} + o_p(n_0^{-\frac{1}{2}}) \\
= & U_{N,N} + \frac{1}{n_0}\sum_{i\in\tilde{s}_0}\{\phi_N^*(Y_i) - \frac{1}{N_0}\sum_{k\in s_0}\phi_N^*(Y_k)\} + o_p(n_0^{-\frac{1}{2}}),
\end{aligned}
$$

where $\phi_N^*(Y_i) = \nu^2\phi_{n_0,N_0}(Y_i) + \nu(1-\nu)\phi_{n_0,N_1}(Y_i)$. This concludes the first part of Lemma 1. By the asymptotic convergence result of finite population sampling, conditional on $\mathcal{F}_0$,

$$
\frac{1}{n_0}\sum_{i\in\tilde{s}_0}\{\phi_N^*(Y_i) - \frac{1}{N_0}\sum_{k\in s_0}\phi_N^*(Y_k)\} \to 0
$$

in probability as $n_0/N_0 \to \alpha$. This result implies the second part of Lemma 1.

Proof of the convergence result in (6). First, we may use Lemma 1 with $\phi(Y_i,Y_j) = w_{ij}(\theta_0)\dot{\eta}_{ij}(\theta_0)d_{ij}I\{X_j \geq t\}/G^2(X_j)$, and the Strong Law of Large Numbers for U-statistics to show that $q_N(t)$ converges to $q(t)$ almost surely at given t. Moreover, $q_N(t)$ converges to $q(t)$ uniformly on $[0,t_0]$ because $q_N(t)$ is a nonincreasing function in t on a finite interval $[0,t_0]$, and it is bounded by conditions A, C and F(i). Let $\pi_N(t)$ be the full cohort counterpart of $\pi_n(t)$, i.e., $\pi_N(t) = N^{-1}\sum_{l=1}^N I(X_l \geq t)$. By the Gilvenko-Cantelli theorem, $\pi_N(t)$ converges uniformly to $\pi(t)$ on $[0,t_0]$. At given time point t, $\pi_n(t)$ converges to $\pi_N(t)$ by asymptotic results of finite population sampling. Also, $\pi_n(t)$ is monotonic and bounded, so $\pi_n(t)$ converges to $\pi_N(t)$ uniformly in t. This implies that $\pi_n(t)$ converges uniformly to $\pi(t)$ on $[0,t_0]$. Thus, as $n,N\to\infty$, $\sup_{0\leq t\leq t_0}\|q_N(t)/\pi_n(t) - q(t)/\pi(t)\| \xrightarrow{p} 0$ due to the uniform convergence of $q_N(t)$ and $\pi_n(t)$, and the boundness of $\pi(t)$ and $q_N(t)$.

Under condition F(i), it follows from the Martingale Central Limit Theorem that the martingale process $N^{-1/2}\sum_{k\in s}M_k^c(t)$ based on the full cohort data converges weakly to a tight zero-mean Gaussian process. It also follows from Hájek's Central Limit Theorem for finite population sampling, and arguments about tightness in Example 3.6.14 of van der Varrt and Wellner (1996), that the subcohort quantity $n^{-1/2}\sum_{k\in\tilde{s}}M_k^c(t)$ also converges weakly to a tight Gaussian process. Write $B_n(t) = n^{-1/2}\sum_{k\in\tilde{s}}M_k^c(t), t\in[0,t_0]$. The tightness of $B_n(t)$ implies that (Sen and Singer (1993, p.330)) for any $\epsilon,\eta>0$, $pr\{W_\delta(B_n)>\epsilon\}<\eta$

for every $\delta > 0$ as $n \to \infty$, where $W_\delta(B_n) = \sup\{|B_n(t) - B_n(u)| : 0 \le u < t < u + \delta \le t_0\}$. Suppose we partition $[0, t_0]$ with $0 = p_0 < p_1 < \cdots < p_h = \tau$ where $h\delta^* = t_0 < \infty$, and write $H_N(t) = q_N(t)/\pi_n(t)$ and $H(t) = q(t)/\pi(t)$. Then

$$\begin{aligned}
&\int_0^{t_0} \{H_N(t) - H(t)\} dB_n(t) \\
&= \sum_{i=0}^{h-1} \int_{p_i}^{p_{i+1}} \{H_N(t) - H(t)\} d\{B_n(t) - B_n(p_i)\} \\
&= \sum_{i=0}^{h-1} [\{H_N(t) - H(t)\}\{B_n(t) - B_n(p_i)\}]_{p_i}^{p_{i+1}} \\
&\quad - \sum_{i=0}^{h-1} \int_{p_i}^{p_{i+1}} \{B_n(t) - B_n(p_i)\} d\{H_N(t) - H(t)\} \\
&= (I) + (II), \text{say.}
\end{aligned}$$

Note that $\max_{0 \le i \le h-1} |B_n(p_{i+1}) - B_n(p_i)| \le W_{\delta^*}(B_n)$, where for $n \ge n_0(\epsilon, \eta)$, $pr\{W_{\delta^*}(B_n) \le \epsilon\} > 1 - \eta$. Also, $\max_{0 \le i \le h} \|H_N(p_i) - H(p_i)\| \le \sup_{t \in [0, t_0]} \|H_N(t) - H(t)\| \xrightarrow{p} 0$. Thus, term (I) $\xrightarrow{p} 0$ as $n \to \infty$. As previously stated, $q_N(t)$, $\pi_n(t)$, $q(t)$ and $\pi(t)$ are all nonincreasing and of bounded variation on $[0, t_0]$. Hence, term (II) can be shown to converge to zero in probability as $n \to \infty$ by the Dominated Convergence Theorem.

Proof of Lemma 2. Self and Prentice (1988) gave a proposition similar to this lemma. We prove it using a different approach. Note that given $\mathcal{F}$, $G_N(X)$ is fixed while the distribution of $H_N(X, \xi)$ is generated by the $\binom{N}{n}$ equally likely choices of ξ. Thus, the characteristic function $\phi(\lambda_1, \lambda_2)$ of the vector $\{N^{1/2} G_N(X), N^{1/2} H_N(X, \xi)\}$ is

$$E\{e^{it\lambda_1 N^{\frac{1}{2}} G_N(X)} e^{it\lambda_2 N^{\frac{1}{2}} H_N(X,\xi)}\} = E\Big\{e^{it\lambda_1 N^{\frac{1}{2}} G_N(X)} E[e^{it\lambda_2 N^{\frac{1}{2}} H_N(X,\xi)} | \mathcal{F}]\Big\}.$$

It follows from Hájek's Central Limit Theorem for finite population sampling that, conditional on $\mathcal{F}$, $n^{1/2} H_N(X, \xi)$ converges to a normal distribution $N\{0, (1-\alpha)\sigma_f^2\}$. Thus,

$$\phi(\lambda_1, \lambda_2) \approx e^{\frac{1}{2}(\lambda_2 t)^2 (1-\alpha)\alpha^{-1} \sigma_f^2} E\{e^{it\lambda_1 N^{\frac{1}{2}} G_N(X)}\} \approx e^{\frac{1}{2} t^2 [\lambda_2^2 (1-\alpha)\alpha^{-1}\sigma_f^2 + \lambda_1^2 \sigma_g^2]},$$

where the approximate sign indicates asymptotic equivalence. Since the right hand side represents the characteristic function of a bivariate normal distribution with null mean vector and a diagonal dispersion matrix as defined in (2), the proof is complete.

References

Anscombe, F. J. (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc.* **48**, 600-607.

Cheng, S. C., Wei, L. J. and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835-845.

Cheng, S. C., Wei, L. J. and Ying, Z. (1997). Predicting survival probabilities with semiparametric transformation models. *J. Amer. Statist. Assoc.* **92**, 227-235.

Fine, J., Ying, Z. and Wei, L. J. (1998). On the linear transformation model for censored data. *Biometrika* **85**, 980-986.

Fleming, T. R. and Harrington, D. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Foutz, R. V. (1977). On the unique consistent solution to the likelihood equations. *J. Amer. Statist. Assoc.* **72**, 147-148.

Gill, R. D. (1980). *Censoring and Stochastic Integrals*. Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publ. Math. Inst. Hungar. Acad. Sci.* **5**, 361-374.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47**, 663-685.

Kong, L., Cai, J. and Sen, P. K. (2004). Weighted estimating equations for semiparametric transformation models with censored data from a case-cohort design. *Biometrika* **91**, 305-319.

Kulich, M. and Lin, D. Y. (2000). Additive hazards regression for case-cohort studies. *Biometrika* **87**, 73-87.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1-11.

Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Statist.* **16**, 64-81.

Sen, P. K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference*. Wiley, New York.

Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics*. Chapman and Hall, New York.

Sherman, R. P. (1994). U-Processes in the analysis of a generalized semiparametric regression estimators. *Econom. Theory* **10**, 372-395.

Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

313 Parran Hall, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, 130 DeSoto Street, Pittsburgh PA 15261, U.S.A.

E-mail: lkong@pitt.edu

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, U.S.A.

E-mail: cai@bios.unc.edu

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, U.S.A.

E-mail: pksen@bios.unc.edu

(Received June 2004; accepted September 2004)